2 État de l'art et démarche retenue

Les langues des signes, par la quadridimensionnalité dans laquelle elles s'inscrivent et la simultanéité d'information dont elles procèdent, sont difficilement analysables. Les systèmes de transcription, le plus souvent manuels et mis au point par les chercheurs en fonction des phénomènes linguistiques qu'ils souhaitent étudier, se révèlent peu satisfaisants.
En préalable à un premier travail de transcription réalisé par les membres de chaque laboratoire participant au projet LS-COLIN afin de définir un cadre de recherche commun, un récapitulatif des différents systèmes de notation existants a été fait.

2.1 Relevé des différents systèmes d'écriture et de transcription existants

2.1.1 Les systèmes de notation manuels

- **Roch-Ambroise Bébian**. La plus ancienne tentative connue de notation d'une langue des signes date du XIXe siècle. En 1825, **Roch-Ambroise Bébian**, professeur français spécialisé dans l'enseignement des Sourds, écrit 'Mimographie', ouvrage dans lequel il propose une description de la LSF en moins de 200 symboles, en prenant en compte la mimique faciale, la description de la forme des mains etc.… Ce système, pourtant précurseur, n'a pas été exploité, un congrès international réunissant des partisans de l'oralisme s'étant tenu à Milan en 1880 et ayant préconisé, pour des raisons essentiellement politiques et religieuses, l'interdiction de la langue des signes dans les milieux éducatifs. Cela mit fin, pour des décennies, à toute recherche sur les langue des signes, et, partant, à toute tentative de transcription de ces langues.
- **William Stokoe**. Ce n'est qu'au début des années 1960 que s'amorce un regain d'intérêt pour les langues des signes et qu'on voit réapparaître, aux Etats-Unis, un nouveau système d'écriture, mis au point par le linguiste **William Stokoe**. Ce dernier, dont le but était de démontrer la double articulation de l'ASL (American Sign Language) afin de la faire accéder au rang de langue du monde, évacue l'iconicité de ses recherches car considérée par lui comme non pertinente et appelée à disparaître. Il établit, en se basant sur les travaux de Bébian, que les signes ne sont pas des gestes holistiques, mais sont décomposables en une configuration, une orientation, un emplacement et un mouvement, paramètres auxquels il donne le nom de 'chérèmes', équivalents gestuels des phonèmes des langues orales. Ce système de transcription élaboré par Stokoe comporte 55 symboles représentant 55 chérèmes, et sera utilisé pour la réalisation d'un dictionnaire paru en 1965, comportant plus de 3000 entrées. Toutefois, cette transcription phonologique des signifiants de l'ASL présente l'inconvénient de ne noter que les gestes, car on pensait à cette époque que l'essentiel des effets de sens ne passaient que par les signes. Outre le fait de rejeter l'iconicité 'hors langue', et donc de passer à côté d'une des caractéristiques constitutives des langues des signes, il ne prend pas en compte la mimique faciale, et laisse de côté le regard, paramètres pourtant déterminants pour la construction du sens.
- **Paul Jouison**. Éducateur spécialisé de formation, Paul Jouison rencontre le monde des Sourds au début des années 70, et met au point un système d'écriture de la LSF, baptisé D'SIGN (le D est là pour « discours »), qui n'a pas pour but de décrire la forme visuelle des signes, mais de restituer leur dynamique de production. Jouison se place en effet dans une perspective énonciative et discursive. Il distingue d'abord l'orientation du regard et l'action du visage, et souligne ensuite l'importance de la localisation (ou orientation) du signe par rapport au corps du signeur. Jouison élabore aussi un classement des configurations manuelles selon une logique interne, qui les rend iconique, auxquelles il

faut ajouter un élément supplémentaire, qui correspond à peu près aux classificateurs de formes (proformes). Enfin, le dernier paramètre retenu est le mouvement global du corps qui, comme le regard, aura un rôle important dans la cohésion de l'énoncé. Globalement, ce système témoigne d'une grande précision et d'une observation minutieuse des paramètres gestuels. Mais sa complexité le rend difficilement utilisable par un chercheur non averti, et rend peu gérable la transcription de séquences entières de récits.

- **Le système HamNoSys** (Hambourg Notation System). Élaboré en Allemagne par un groupe de chercheurs sourds et entendants, commencé au début des années 70, il reprend les quatre paramètres retenus par Stokoe, auxquels il ajoute la mimique faciale. Il se distingue du système précédent car il adopte une démarche plus phonétique (description de l'ensemble des phénomènes concourant à la réalisation d'un signe) que phonologique (approche distinctive visant à mettre en évidence un inventaire de chérèmes). De plus, le classement des configurations de la main obéit à des critères fonctionnels et articulatoires. Ce système est iconique, alors que celui de Stokoe est surtout dactylologique, et les caractères de transcription d'HamNoSys, idéographiques, permettent un décodage rapide. Un logiciel informatique, regroupant tous les symboles, existe depuis 1989. Mais leur trop grand nombre (on recense environ 500 symboles) rend leur mémorisation très difficile et ce système ne satisfait pas, en définitive, des conditions optimales de lisibilité.
- **La transcription en partition**. Elle se met en place au début des années 80, et est aujourd'hui largement utilisée par la communauté linguistique internationale. Elle vise l'analyse du discours et fait apparaître la simultanéité ainsi que la corrélation structurelle entre les paramètres de chaque signe. Les explications relatives aux différents paramètres sont données dans la langue du chercheur et en désignent, selon les paramètres considérés, la face signifiante (par exemple pour le regard, les mouvements du visage, mouvements de la bouche…) ou la face signifiée (ainsi pour la mimique faciale, dans la mesure où il est impossible d'établir un système lisible qui rendrait compte de tous les muscles du visage entrant en jeu pour telle ou telle expression). En ce qui concerne les signes standard, ils sont écrits entre crochets, le mot inscrit ne devant en aucun cas être associé au signifié du signe, ni même véritablement à son sens, mais plutôt à l'invariant signifiant de ce signe, auquel on associe une fois pour toute tel mot de la langue orale considérée.

2.1.2 Les systèmes de notation informatisés

Les années 90 voient l'émergence, dans divers pays, de différents systèmes informatisés de transcription et d'écriture des langues des signes.

- **Sync Writer.** La première version de ce programme allemand de transcription pour MAC date de 1991. La seconde version, plus élaborée, paraît en 2000. Elle utilise le procédé de transcription en partition en adoptant le champ classique des classifications linguistiques, et intègre le système HamNoSys. Doté de pistes configurables, ce programme permet l'inclusion d'un support d'images au moyen de Quick Time.
- **Sign stream.** Ce programme, qui s'inscrit dans un vaste projet de recherches linguistiques sur l'ASL entrepris par les universités de Boston et de Gallaudet (Washington), constitue un outil de gestion pour une base de données informatisée complète des signes de cette langue. Il reprend le concept de transcription du paradigme gestuel sous forme de partition se déroulant dans le temps, et présente aussi l'intérêt d'intégrer des images vidéo numérisées au-dessus de chaque segment transcrit. Toutefois, ce système est incomplet car il ne prend en compte ni la mimique faciale, ni la direction du regard dans l'analyse.
- **Sign Writing**. Ce système propose une 'écriture' des langues des signes. Il est issu, au départ (1974), du Dance Writing, et s'inscrit dans un système plus général de notation et d'écriture du mouvement (Sutton Mouvement Writing & Shortland). Utilisé essentiellement, lors de sa création, pour la transcription d'enregistrements vidéo destinés

à la recherche, ce système a beaucoup évolué et il existe aujourd'hui plusieurs axes d'application, dont on retiendra : a) le développement d'une forme d'écriture des Langues des Signes, applicable dans l'enseignement de ces langues (tant en primaire que dans le secondaire) ; b) l'écriture et la lecture de contes en Langues des Signes ; c) la préservation du registre théâtral des LS sous forme écrite.

- **SignPS**. Issu du projet européen TIDE (Technology Initiative for Disabled and Elderly People-1994-1997), ce moyen de notation, utilisable par les chercheurs et par le grand public, possède son codage propre, et permet la reconstitution de signes isolés.
- **Sign Synthesis**. Inspiré de l'analyse de la synthèse de la parole, le projet SignSynth est développé par le Département de Linguistique de l'Université du Nouveau-Mexique. C'est un système d'analyse phonologique qui associe à une image de synthèse les paramètres (configuration, orientation, emplacement et mouvement) étudiés par Stokoe.
- **Kheiros**. Conçu par Alexandre Bonucci (université de Lyon 2), dans le cadre de sa thèse de doctorat, Kheiros constitue une base de données d'entrées lexicales de la LSF, basée sur une analyse phonologique et une indexation figurative des signes.

2.2 Les transcriptions réalisées dans le cadre de LS-COLIN

Une fois présenté cet état des lieux des systèmes existants, nous avons décidé de soumettre aux participants ayant déjà réalisé, dans le cadre de leurs travaux, la transcription d'une langue des signes, un court passage du conte Blanche Neige, signé en LSF, à transcrire, afin d'appréhender les intérêts de chacun et de définir les possibilités d'un travail commun, et les bases sur lesquelles il devait être mené.
Ce travail a été réalisé entre décembre 2000 et mars 2001. La séquence à transcrire, la même pour tous, durait 17 secondes environ. On n'en reportera ici qu'une infime partie, afin de mettre en exergue les préoccupations de chacun et les critères de segmentation retenus.

Transcription n°1

Fanch Lejeune (doctorant en co-direction LIMSI (Orsay) et Paris IV).

Situation	Situation 1	
Nature de la situation	Mouvement par rapport à un repère : " descendre d'un cheval "	
Informations sur les entités	md : locatif - config 'main plate' -> cheval MD : config 'V'- " jambes " -> prince	
Phases saillantes de la situation		
Description de ces phases	main dominante sur main dominée	main dominante à coté main dominée
Information cinématique	mouvement vers le bas de la main dominante depuis main dominée	
Informations dynamiques	mimique : 'soutenue' / mouvement du buste vers la gauche	
Autres informations	Regard :' devant'	

Transcription n°2

Annie Risler (post-doctorante. Paris 8)

	ALORS	JE DESCENDS DE CHEVAL
Corps	mouvt tête	mouvt des épaules qui accompagne les mains
Regard	Vague	vague, devant
mimique	Résolu	contrôle d'action
Mains		D descend de G
D		class "jambs"
G		class. base vertical

Transcription n°3

Gwénaëlle Jirou (DEA, Paris 8)

N° séquence	1	2
Durée de la séquence	02 :56	00 :68
SIGNES	descendre (de cheval)	marcher
Main dominante	personnage -----------------------------------	jambes --- ---------------------
Main dominée	TF : cheval (locatif) --- ---	
2 mains		
Transferts	TS --- -----	
Regard	vers mains	vers caméra--------------------------------------- ---------------------
Mvts de tête	en arrière	face --- ---------------------
Mimique faciale	Assertive	décidée (lèvres serrées et étirées)
Mvts du corps	haussement d'épaules figurant l'effort	balancement de la marche -------------------- ---------------------
Traduction en Français	*Le personnage (le prince) descend de cheval. Il marche.*	

<u>Segment 1</u> : 03'' :24''' <u>Segment 2</u> : 03'' :92'''

Transcription n°4

Marie-Anne Sallandre (Doctorante, Paris 8)

Regard	face ___________________________________
- Do	(2 jambes) descendre du cheval - faire qq pas
Signes manuels – 2	
- Dé	locatif: cheval ____________________________
type de transferts	TS ____________________________________
mimique faciale	bouche plissée: intriguée et résolue _____________
orientation du corps	face ___________________________________
Traduction française	Poussé par la curiosité, il descendit de cheval

Transcription n°5

Ivani Fusellier-Souza (Doctorante Paris 8)

Fragment : 1	1c	1d	1f
Images Direction des mouvements			
Regard	________________	Vers le locatif	Vers la camera
Main dominante	Début de transfert de forme : " emplacement des jambes du personnage "	TS: début du déplac. De líactant	TS : déplac. De líactant vers le sol.
Deux mains	.	.	.
Main dominée	Début du locatif (Cheval)	Locatif : cheval	Locatif : cheval
Mimique faciale	Résultative, prise de décision	résultative, décidé	résultative, décidé
Mouvement de la tête	Penchée complètement en arrière	Redressé	En face
Mouvement de la bouche	Lèvres serrées et contractées [mm]	______________	______________
Mouvement de la partie supérieure du corps	Légèrement penché en arrière	Droit, en face	Balancement des épaules vers la droite
Traduction approchée	*descendre de son cheval.*		

Transcription °6

Nathalie Monteillard (DEA, Paris 8)

	b
Regard	Vers MD puis vers un point sur la gauche (endroit d'où provient la voix)
MD	**descendre de cheval**
2M	
Md	**locatif** (cheval)
Mimique faciale	espiègle
Mouvement de la tête	revient de face
Mouvement de la bouche	lèvres pincées et étirées (rictus à G)
Mouvement de la partie supérieure du corps	aucun
Dynamique Enonciative	Double Transfert

Traduction : (**b**) Il descend de cheval.

Transcription n°7

Annelies Braffort (Maître de Conférences, LIMSI) : transcription des configurations
Christophe Collet (Maître de Conférences, LIMSI) : transcription du regard
(Remarques préalables : nos transcriptions habituelles comportent des données numériques fournies par nos systèmes de capture (gant ou caméra) ainsi que des données symboliques associées à ces données, que nous ne pouvons donc pas fournir sur la séquence Blanche-Neige).

Les parties A et B contiennent les transcriptions de type symbolique relatives aux configurations des mains dominante et dominée. La partie C contient la transcription de type symbolique relative au regard.
Nous avons segmenté la séquence vidéo en utilisant les numéros d'images.

A. Transcription des configurations

1. Mode d'emploi de la transcription *:*

Chaque ligne définit une zone de la séquence d'images durant laquelle la configuration est constante. Elle est décrite par la séquence : " description de la configuration " " numéro d'image"

- La description de la configuration est donnée par un terme classique (pince, main plate…), un chiffre ou une lettre (5, s…) ou un terme lié au contexte (jambe, jambe bouge).
- Le numéro est celui de la 1ère image de zone.

Notation :

- Transition = zone de co-articulation entre 2 zones pour lesquelles la configuration est constante

Liste des descriptions utilisées :

- pince, jambe, jambe bouge, main plate, angle, s, n, 5, o, index, 2, neutre

  Exemple :

```
• pince 1
transition 2
jambe 39
...
```

  Lecture de l'exemple :

« On observe une configuration " pince " dans l'image 1, puis une configuration de " transition " de l'image 2 à l'image 38, puis une configuration " jambe " de l'image 39 à l'image ... »

2. Transcription pour la main dominante

Pince	1
Transition	2
Jambe	39
jambe bouge	87
Transition	109
main plate	116
Transition	123

3. Transcription pour la main dominée

Neutre	1
Transition	31
main plate	39
Transition	123

B. Transcription du rapport entre les deux mains

1. Mode d'emploi de la transcription

Chaque ligne, sauf la dernière, définit une zone de la séquence d'images durant laquelle le rapport entre les deux mains est constant.
Elle est décrite par la séquence : " description du rapport " " numéro d'image ".

- La description du rapport est donnée par une périphrase (md glisse sur MD).
- Le numéro est celui de la 1ère image de zone.

Notations :

`md` = main dominée
`MD` = main dominante
`transition` = zone de co-articulation, entre 2 zones pour lesquelles le rapport entre les 2 mains est constant

Liste des descriptions utilisées :

- 1 main (si une seule main), md glisse sur bras MD, 2 mains symétriques, MD au dessus md, 2 mains se touchent, 2 mains asymétriques

  Exemple :

```
1 main 1
transition 32
md glisse sur bras MD 43
...
```

Lecture de l'exemple :

« On observe une seule main de l'image 1 à l'image 31, puis une "transition" de l'image 32 à l'image 42, puis la md glisse sur le bras de la MD de l'image 43 à l'image ... »

2. Transcription

1 main	1
transition	32
md glisse sur bras MD	43
transition	102
2 mains symétriques	123

C. Transcription du regard

1. Mode d'emploi de la transcription

Chaque ligne, définit une zone de la séquence d'images durant laquelle le regard est constant. Elle est décrite par la séquence : " description du regard " " couple de 2 numéros d'image ".

- La description du regard représente soit la direction (gauche, face, ...), soit le fait que les yeux sont fermés (ferme).
- Les numéros d'images décrivent un intervalle (1ère image de la zone ; dernière image de la zone)

Liste des descriptions utilisées :

- Gauche, ferme, face, bas-gauche, face-plisse

Exemple :

```
Gauche 1-6
ferme 7-46
face 47-51
```

Lecture de l'exemple :

« La narratrice regarde à gauche de l'image 1 à l'image 6, puis elle ferme les yeux de l'image 7 à l'image 46, regarde en face de l'image 47 à l'image 51, ... »

2. Transcription

Gauche	1-6
ferme	7-46
face	47-51
Gauche	52-54
Face	55-58
Gauche	59-59
Bas-Gauche	60-61
Face	62-101
ferme	102-103
Face	104-109
ferme	110-114
face	115-172

2.3 Conclusion

La confrontation de ces différents modes de transcription attestait que, si les linguistes avaient tous choisi une édition en portée, en revanche le versant informatique exposait une analyse notoirement différente. De plus, les différentes partitions révélaient des voies différentes de prise d'indices. Il était par conséquent difficile, étant donnée l'hétérogénéité de ces différentes versions, de s'entendre et de trouver une base commune de recherche, à moins d'envisager la mise au point d'un éditeur de partition qui servirait aux linguistes à l'analyse des langues de signes, et constituerait, pour les informaticiens, un outil pour la communication homme/machine.

Une réunion était donc nécessaire afin d'établir conjointement une grille d'analyse, qui servirait de base aux informaticiens pour la création de cet éditeur.

2.4 Grille d'analyse (valable pour un monologue)

Niveaux			**Définitions**	**Indices de segmentation**
Discours			interaction/thème	changement de thème
Enoncés			cohérence sémantique et syntaxique	changement d'actant, de plan, suppression d'actant, clignement de paupières, cht direction regard
Mimique			point de vue du signeur/discours	croisement regard/ « oui » x3 dictum (le fait) / modus (comment c'est dit)
unités signifiantes	Signe	**Syntagme**	cohérence syntaxique d'un groupe de morphèmes	clignement de paupières (pas toujours) pic de tension musculaire sur noyau prédicatif
		Morphème	pas de cohérence syntaxique : signes standards	pauses
	paramètres	**Morphèmes liés**	niveaux des paramètres. ex: emplacement tête = activité cérébrale	
unités non signifiantes	**Phonétique**		description articulatoire des paramètres	modification de la configuration, changement d'emplacement
	Prosodie		rythme: long, bref, répétitions, tension	

3 Conception et élaboration du corpus

L'objectif était de réaliser un corpus qui permette pour les linguistes de mettre en évidence l'iconicité de la langue des Signes (grammaire spatiale) et d'apporter un support de haute qualité pour les informaticiens en traitement de d'image.
La LSF étant une langue naturelle et non un artefact, il fallait donc définir des énoncés avec des structures différentes pour faire émerger la langue dans ses multiples composantes. Nous avons donc varié les genres discursifs afin de constituer un éventail des langages possibles : récits complets à partir d'images (histoire du *Cheval,* de l'*Oiseau*), discours en interaction, argumentation, etc.
La production de ces énoncés par des informateurs d'âge, sexe, profession et origine régionale très différents et une grande liberté dans leurs productions étaient un gage de qualité du contenu de ce recueil. Ceci a impliqué de la part des chercheurs un échange constant avec des locuteurs natifs de la langue des signes participant activement à cette réalisation.
Dans le cadre d'une convention avec l'INJS (Institut National des Jeunes Sourds) de Paris, l'enregistrement, sous la maîtrise des techniciens du studio Photo et Vidéo (eux-mêmes pratiquant la LSF), réunissait les meilleures conditions d'accueil et de liberté d'expression pour nos informateurs.
L'enregistrement sur vidéo numérique permettant un arrêt sur image précis et assurant une bonne portabilité était le meilleur support actuel pour la réalisation de ce corpus.
Les séries de critères définis pour être étudiés (orientation du regard, position du corps, de la tête et des épaules, mimique faciale…) nous ont conduits, après plusieurs essais, à un enregistrement avec 3 caméras qui permet la prise en compte simultanée de tous ces paramètres linguistiques.
Un montage des trois vues différentes sur le même film est en cours de réalisation.
Ce corpus vidéo sera transféré sur des supports différents : VHS, CD et DVD, en fonction des utilisateurs potentiels.

3.1 Spécifications de réalisation

3.1.1 Les locuteurs :

Le recrutement de locuteurs sourds adultes présentant une variété sociolinguistique maximale et une bonne compétence en langues des signes a eu lieu par différents moyens (mailing, fax, SMS, …) et supposait une bonne connaissance de la communauté des Sourds français. Il a été effectué sur une période de trois mois par les deux doctorantes en linguistique de la LSF de Paris 8 qui ont contacté environ 35 personnes. Finalement, 13 personnes ont répondu positivement. Le temps dévolu à l'établissement des dossiers pour le défraiement de ces personnes n'est pas non plus à négliger. En effet, nous avons constaté que le défraiement de locuteurs participant à un enregistrement audio ou vidéo n'est actuellement pas une pratique institutionnalisée (et donc facilitée) dans les Universités françaises. Une fiche d'engagement (*Annexe 1*) nous permettant de nous assurer du droit des images sur les films produits a été signée par chaque intervenant, ce qui nous met en règle vis-à-vis de la loi sur la propriété intellectuelle et nous autorise à reproduire et diffuser notre corpus dans un cadre précis.

3.1.2 Les techniques d'acquisition :

Des essais d'enregistrement ont été effectués à l'INJS le 26 octobre 2001. Le but était de déterminer les conditions idéales d'éclairage, de positions de caméra qui devaient convenir à

la fois aux linguistes pour la compréhension de la LSF et aux informaticiens pour l'analyse des images. Nous en avons retenu la configuration suivante :

- 3 caméras numériques professionnelles, une de face en plan américain, une de face en contre plongée située à 2 m devant le locuteur et à 0m75 de hauteur, la dernière au-dessus du locuteur.
- Eclairage maximum : 6 projecteurs avec des réflecteurs de part et d'autre du locuteur permettant de minimiser au maximum les ombres.
- Un fond bleu, sur lequel les locuteurs portant un vêtement noir à manches longues viendraient se détacher.
- La synchronisation des 3 caméras serait assurée par un flash.

3.2 Réalisation

L'enregistrement a eu lieu les 11 et 12 janvier 2002 à l'INJS de Paris avec les personnes suivantes :

- les techniciens du studio Photo et Vidéo : Stéphane MANGAUD et Grégory GONZALEZ.
- En présence des équipes participant au projet :
  - Pour les Sciences du langage de l'Université Paris 8 : Marie-Anne SALLANDRE, Ivani FUSELLIER et Gwenaëlle JIROU (vendredi)
  - Pour le LIMSI : Annick CHOISIER, Christophe COLLET (vendredi) et Fanch LEJEUNE (vendredi après-midi)
  - Pour L'IRIT : Boris LENSEIGNE

3.2.1 Ces journées ont été organisées par :

Marie-Anne SALLANDRE et Ivani FUSELLIER du département des Sciences du Langage de l'Université Paris 8 pour le recrutement des signeurs, l'élaboration des consignes et du protocole expérimental, et l'établissement des dossiers pour le paiement des signeurs.
Annick CHOISIER et Christophe COLLET du LIMSI pour les aspects matériels : achat et gestion des différents formats de cassettes vidéo, convention avec l'INJS, droits des images…

3.2.2 Conditions techniques :

Nous avons adopté l'option maximale des conditions définies le 26 octobre 2001, c'est-à-dire 3 caméras, un fond bleu et un très fort éclairage (voir *photo 1*).
La caméra 1 une SONY DVCAM située à 5 m face au signeur, et à une hauteur de 1m60 filmait la personne en plan américain,
La caméra 2 une SONY mini DVCAM située à 2m55 face au signeur et à 0m90 en hauteur cadrait la tête en contre-plongée.
La caméra 3, une mini DV au-dessus du signeur, hauteur depuis de sol : 2m60 enregistrait l'amplitude des mouvements de bras et de buste.
Un flash, au début de l'enregistrement devait permettre de synchroniser les images des 3 caméras, au montage.



Photo 1 : préparation du tournage

Le protocole était défini comme suit : chaque signeur, convoqué à une heure précise était accueilli par l'une des trois linguistes qui lui expliquait (voir *photo 2*), dans une salle à part, ce qu'on attendait de lui, à savoir :

1. Raconter l'histoire en images du *Cheval*.
2. Raconter l'histoire en images de l'*Oiseau*.
3. S'exprimer sur l'un des deux sujets (au choix, ou les deux, suivant les locuteurs): le passage à l'euro ou les évènements du 11 septembre 2001.
4. Expliquer sa recette de cuisine préférée.
5. Pour les personnes ayant suivi le cursus de linguistique de Paris 8 (diplôme du DPCU), choisir un thème du programme de linguistique et l'expliquer à la manière d'un cours.
6. Raconter à nouveau l'histoire du *Cheval*.



Photo 2 : explication des consignes par le linguiste
(Ivani Fusellier, à droite) s'adressant en LSF au locuteur (Juliette Dalle, à gauche)

Les intervenants préparaient leur prestation pendant un temps moyen de 20 minutes puis venaient dans le studio d'enregistrement pour être filmés, face aux caméras et face à la personne qui les avaient pris en charge, ils devaient s'adresser à elle.
La durée de la production varie entre 5 à 8 séquences consécutives par locuteur, en fonction de sa disponibilité, de sa mémorisation et de ses envies. Le but étant de rester le plus naturel possible, nous sommes intervenus au minimum lors de la prise de vue. Nous obtenons un total

de 90 productions différentes (dont récits, explications techniques, explications métalinguistiques, argumentation) *(Annexe 2)*.
Des interruptions à cause d'incidents techniques (batterie déchargée, ...) ou sur demande de l'intervenant ont pu se produire pour certains locuteurs ; dans ce cas, nous avons essayé de reprendre toute la série depuis le début.
13 locuteurs ont ainsi participé à la réalisation de ce corpus, nous avons donc environ 6 heures de film à analyser (2 heures par caméra).

Le détail du corpus brut est donné en annexe (tableau de synthèse + liste détaillée des productions).

3.3 Evaluation

La qualité technique de ces films numérisés, qui nécessite des moyens informatiques importants, est validée par le confort de visualisation qu'elle procure. Les montages, transferts et productions finalisées nous ont demandé d'acquérir des techniques lourdes de traitement de films.
Bien que l'analyse linguistique ne soit pas achevée, nous constatons que le panel obtenu des productions en LSF est très satisfaisant et est actuellement inédit en France, ainsi que sur le plan international. Au cours du projet, un corpus de Langue des Signes Primaire (LSP) et de LIBRAS (Langue des Signes Brésilienne) a également été réalisé, au Brésil, dans le but d'effectuer une analyse transversale d'un même récit (l'histoire du *Cheval*).
La mise à disposition de ces données vers des communautés scientifiques différentes (linguistes, informaticiens, pédagogues sourds de la LSF, …) reste un des buts à atteindre. Ce sont en partie ces communautés qui, par le dépouillement et l'analyse de ce corpus dans leurs différentes spécialités, valideront notre démarche.

4 Réalisation de l'éditeur de partition

4.1 Spécifications

4.1.1 Rôle de l'éditeur

Le rôle de l'éditeur de partition est de permettre la visualisation et la manipulation de séquences vidéo de LSF et la réalisation de transcriptions de vidéos selon des critères utiles à la fois aux linguistes et aux informaticiens. Ces critères doivent donc être paramétrables à chaque utilisation en fonction du but poursuivi ou des informations recherchées dans lors de la réalisation de la transcription (par exemple des structures langagières ou des indices visuels permettant la segmentation de la séquence). Enfin, il fournit un environnement permettant l'application d'opérateurs de traitement d'images sur la vidéo :

- Visualisation et manipulation de séquence vidéo : L'éditeur permet la visualisation de films numérisés, à la manière d'un magnétoscope (lecture, arrêt sur image), ainsi qu'un accès non linéaire aux images de la séquence,
- Transcription de vidéos : Il permet aussi de réaliser une transcription de vidéos, c'est-à-dire d'associer à une ou plusieurs images des informations symboliques ou numériques. Une fois la transcription réalisée, on peut la sauvegarder afin de pouvoir la restituer lors d'un chargement ultérieur de la séquence vidéo.
- Application d'opérateurs de traitement d'image : L'éditeur de partition intègre des fonctionnalités permettant de convertir les données contenues sous forme compressée dans la vidéo dans une forme manipulable par des opérateurs de traitement d'images.

4.1.2 Aspect graphique de l'éditeur

L'éditeur est constitué de deux parties (Figure 1). La partie supérieure contient la visualisation de la séquence vidéo, avec une série d'icônes à cliquer permettant d'accéder rapidement aux fonctionnalités du logiciel et des boutons d'actions sur la vidéo. La partie inférieure comporte la transcription associée.

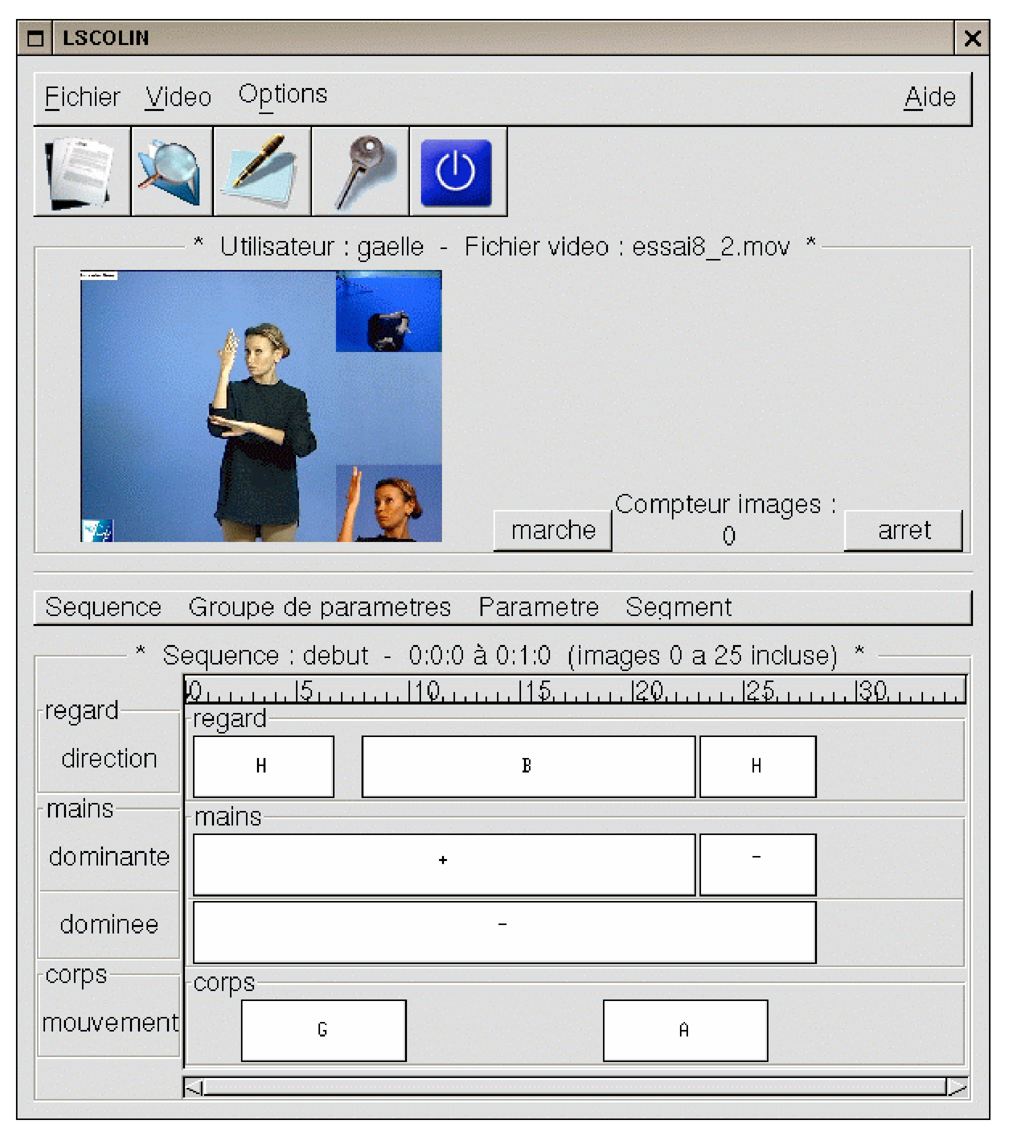


Figure 1 : aspect graphique de l'éditeur

4.1.3 Description détaillée de la partie transcription

La transcription est représentée sous forme d'une partition. L'axe horizontal représente le temps et l'axe vertical contient l'ensemble des paramètres sélectionnés (ex : direction du regard, mouvement des mains, fonction syntaxique, type de transfert, etc…) Ces paramètres sont définis par l'utilisateur en fonction du but poursuivi lors de son analyse.

Sur l'axe horizontal : Chaque paramètre est décrit au sein d'une bande horizontale découpée en sections pouvant elles-même être découpées, le cas échéant en segments :

- La première section à partir de la gauche contient le nom du paramètre (ex : « Signe », « Main Dominée »).
- La section suivante contient une série de segments représentant les valeurs correspondantes du paramètre pour une séquence temporelle donnée (ex : « bonjour »). Lorsque l'on clique sur un segment, un menu contextuel déroulant donne la liste des valeurs disponibles pour le paramètre correspondant et permet d'effacer le segment ou d'accéder à une fenêtre de dialogue pour ajouter une nouvelle valeur. Lorsque l'on clique sur un espace vide de la bande, un nouveau segment apparaît, qu'il est possible de modifier en durée, en emplacement et en contenu grâce au menu contextuel ;

- On peut trouver des bandes de paramètre « fermées », pour lesquelles il y a des segments de valeurs sur toute la bande et des bandes « ouvertes », décrivant des événements à un moment donné de la séquence vidéo.

Sur l'axe vertical :

- On peut regrouper certains paramètres au sein d'un « groupe de paramètres » visuellement identifié lorsque cela est justifié (ex : main dominante + main dominée + deux mains).
- Une barre verticale synchronisée avec le défilement de la vidéo peut se déplacer le long de la transcription selon l'axe temporel.

Une transcription vidéo est composée de plusieurs partitions différentes. En effet, un utilisateur peut vouloir étudier plusieurs extraits d'une même vidéo ou transcrire de plusieurs manières différents la même vidéo.

4.1.4 Généricité, flexibilité, ouverture

L'éditeur peut être utilisé à la fois par des linguistes et par des informaticiens. Il est donc prévu des facilités de personnalisation et de modification de la partie transcription.

Chaque transcription, avec ses séquences, ses paramètres avec leurs différentes valeurs et les segments qui la composent, est décrite et sauvegardée dans un fichier de type texte. Il est possible d'ajouter de nouvelles séquences, de nouveaux groupes et paramètres, ainsi que de nouvelles valeurs pour chaque paramètre et de nouveaux segments. Les valeurs peuvent être représentées selon plusieurs types :

- **I** : icône (ex : « ✋ » pour configuration main plate),
- **C** : symbole (ex : « F » pour regard de Face),
- **G** : description littérale (ex : « bonjour »),
- **V** : valeur numérique (ex : 5cm).

Ainsi, chaque utilisateur peut ajouter :

- de nouveaux paramètres
- de nouvelles valeurs aux paramètres

Chaque utilisateur peut choisir :

- les paramètres qu'il souhaite étudier
- le type d'affichage pour chaque paramètre (icône, code?)
- l'ordre dans lequel ces paramètres sont affichés.

Un utilisateur peut récupérer une transcription déjà réalisée et modifier la forme de la présentation. Pour cela, des profils utilisateurs sont associés à chaque transcription. Ainsi, à la réouverture d'une transcription donnée par un utilisateur donné, celui-ci retrouve la configuration de l'éditeur telle qu'il l'avait définie précédemment.

4.2 Apports du traitement d'images

Les apports du traitement d'images à un outil tel qu'un éditeur de partitions peuvent être considérés sous deux aspects. D'une part il permet de faciliter la tâche du linguiste en automatisant partiellement la réalisation de la transcription (segmentation temporelle, ou remplissage automatique des valeurs affectées au segment) à partir de résultats obtenus par l'application d'opérateurs sur tout ou partie de la séquence vidéo. D'autre part, le traitement d'images donne accès à des informations invisibles directement (occupation de l'espace, mesures de vitesse, modélisation cinématique, etc. …). Enfin, le traitement d'images peut être intégré dans une application plus complexe de construction interactive de concepts (voir Dalle 2001). Le linguiste peut comparer la description qu'il a fournie en termes d'indices visuels avec les résultats proposés par ce système qui lui présente les objets correspondant à la description donnée. Par cette mise en correspondance, le linguiste pourra vérifier la pertinence des indices visuels utilisés pour mettre en évidence des structures particulières de

la langue des Signes. Ce dernier domaine d'application constitue toutefois un sujet de recherche à long terme qui ne rentre pas dans le cadre du projet LSCOLIN, il ne sera donc qu'évoqué à la fin de cette section.

4.2.1 Possibilité d'automatisation partielle par traitement d'images

Les informations reportées dans les différents niveaux de la partition proviennent d'une analyse et d'une interprétation, par le linguiste, de la séquence d'images (détermination des composantes de l'image et analyse des mouvements). Afin de l'aider dans cette analyse, et donc de faciliter l'utilisation de l'éditeur de partitions, nous avons étudié dans quelle mesure le traitement d'image permettrait d'automatiser certaines tâches.
Nous allons présenter les différents problèmes que l'on peut tenter de résoudre par traitement d'image, d'une façon générale, puis spécifier ce qui peut être réalisé dans le cadre limité du projet LS-COLIN et enfin nous présenterons les études menées et les résultats intermédiaires obtenus.

4.2.1.1 Analyse au niveau d'une image

Les paramètres de la LS figurant dans la transcription concernent des éléments du corps, comme la main ou les sourcils dont le TI doit savoir retrouver les correspondants dans l'image. Il s'agit donc de segmenter l'image en zones ayant des propriétés visuelles communes comme la couleur (zones de peau), ou de localiser des configurations de pixels particulières (indices visuels). Ce traitement peut être entièrement automatisé ou au contraire mené en coopération avec l'utilisateur, celui-ci indiquant la zone approximative de l'élément à étudier ou désignant un pixel de cet élément, le système de TI se chargeant ensuite de déterminer les frontières précises de l'élément.
Ces éléments peuvent être composés; on doit donc savoir structurer les composants élémentaires en composants plus complexes (yeux + nez + bouche + joues + front ⇒ visage). Enfin, ces composants sont caractérisés par un ensemble de valeurs intrinsèques (<u>position</u> de la main) ou relationnelles (main <u>à la hauteur de</u> la bouche). Ces valeurs peuvent être déterminées par un traitement (plus ou moins) automatique, en effectuant des mesures directes dans l'image (yeux <u>fermés</u>), ou en reconstruisant l'information 3D associée (<u>direction</u> du regard). Elles concernent les éléments de base (<u>orientation</u> de la main) ou des éléments composés (visage <u>souriant</u>). Il peut s'agir de valeurs numériques dans le cas de mesures directes ou de valeurs symboliques nécessitant une interprétation de plus haut niveau. À partir des mesures effectuées sur les entités 2D ou 3D ou à partir des indices visuels et en exploitant des connaissances a priori (modélisation des composants du corps), on peut identifier les entités (région de couleur peau + taille + forme ⇒ main), en différents niveaux d'abstraction suivant la complexité de l'information introduite dans le programme (région de couleur peau ⇒ main ⇒ main droite ⇒ main dominante ⇒ signe standard).

4.2.1.2 Analyse au niveau de la séquence d'images

La transcription concerne aussi le mouvement des mains ou du corps, éléments qu'il faut donc savoir suivre dans la séquence d'images (c'est-à-dire retrouver dans chaque image de la séquence). Les paramètres portent alors sur une caractérisation de ce mouvement (<u>vitesse</u>, direction, signature de trajectoire, classification en mouvements de type particulier). On retrouve ici aussi la notion de structuration, les mouvements élémentaires pouvant être combinés en des configurations plus complexes (<u>balancement</u>), et la notion de mesures relatives (trajectoires <u>parallèles</u>).

Enfin le mouvement étant fugitif, il est intéressant, pour faciliter son analyse par le linguiste, de pouvoir le matérialiser et de le visualiser en superposition dans l'image (techniques de « réalité augmentée »). Il s'agit donc de construire une image qui enregistre l'historique du mouvement en rendant compte à la fois de son occupation spatiale et de son déroulement temporel.
D'une façon plus générale, il est très pertinent de pouvoir représenter l'espace de signation, d'y faire figurer les référents mis en place au cours du discours et de matérialiser les zones désignées par les pointeurs.

4.2.1.3 Remarque

La reconnaissance d'un élément ou d'une configuration peut souvent être établie à partir de sa signature visuelle dans l'image ou dans une image transformée. Par contre les mesures ou les descriptions faisant référence à l'espace de signation (trajectoires, pointages, détection d'un mouvement vers l'avant) nécessitent une reconstruction partielle 3D. L'image seule ne suffit alors pas car des informations ont été irrémédiablement perdues lors du processus de formation de l'image (projection perspective, occultations). Il faut donc disposer d'informations supplémentaires pour pouvoir reconstruire l'information 3D ; elles peuvent être obtenues par :

- raisonnement (étude des lignes de fuite, de la variation de la taille d'un objet)
- ajout de connaissances supplémentaires (par exemple sur la taille réelle des composants) ou de contraintes sur la géométrie de la scène.
- exploitation de modèles abstraits dont le TI déterminera les paramètres (ex : modèle biomécanique du bras).
- ajout de données supplémentaires :
- en utilisant une seconde image, prise depuis un point de vue différent, et en appliquant les techniques de stéréovision,
- en utilisant plusieurs images dans le temps et en interprétant les mouvements dans les images.

Dans notre cadre de travail actuel (séquence d'images vidéo mono-caméra non calibrée), les possibilités de reconstruction 3D sont très limitées.

4.2.2 Le traitement d'images dans le projet LS-COLIN

4.2.2.1 Problèmes abordés

Les linguistes ont recensé les principaux indices visuels et marqueurs utilisés pour analyser la LS ainsi que leurs valeurs signifiées respectives et le contexte dans lequel elles sont évaluées. Ceci a permis de spécifier les paramètres, leur mode de représentation, les relations pouvant lier différentes parties de la transcription, et enfin de définir ainsi la structure et le contenu d'une partition.
On a vu qu'une partition se présente sous forme de plusieurs lignes découpées en cases par des séparations temporelles (correspondant à des événements spatio-temporels). Chaque ligne correspond à un composant étudié et chaque segment temporel correspond à une valeur de paramètre constante pour ce composant. Pour les composants ne faisant pas intervenir d'information d'ordre sémantique, le TI peut intervenir sur chacun de ces éléments :

- Ajout d'une ligne « images » (en plus de la vidéo) permettant de synthétiser une séquence par une attitude caractéristique du signeur, augmentée de graphismes illustrant le signe (flèches, synthèse de mouvement).
- Segmentation temporelle par détection d'événement, c'est-à-dire par détection des changements de valeurs pour une propriété donnée d'un composant tels que des changement de mouvement (orientation ou vitesse), de configuration ou de relations.

- Ajout de valeurs dans les cases, par reconnaissance (de configurations) ou mesures (direction de mouvement, vitesse, etc…)
- Génération de nouvelles lignes, correspondant à de nouveaux concepts évaluables par analyse d'image (relation entre plusieurs indices)

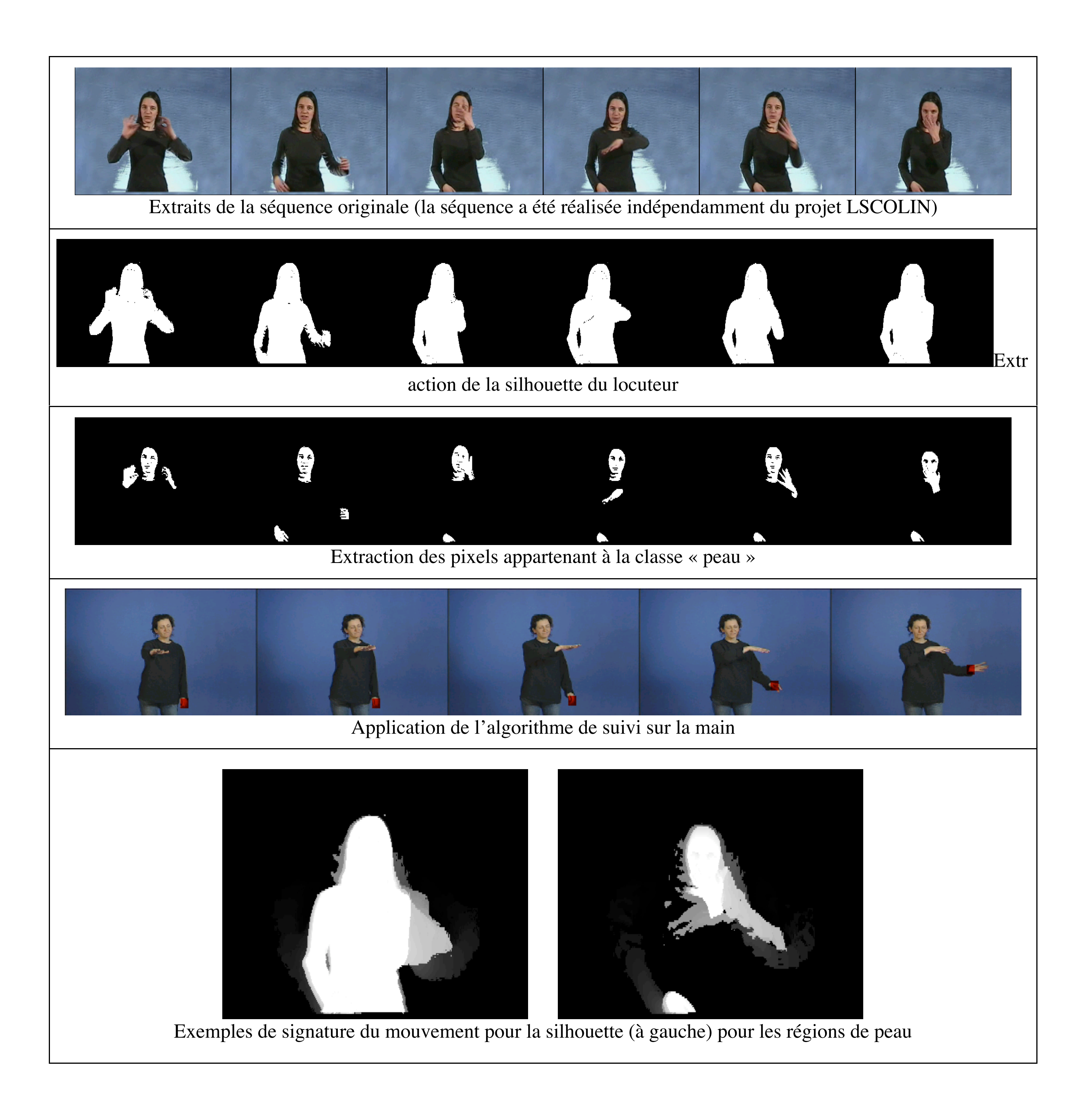
Les travaux en TI ont donc consisté à étudier et à réaliser les outils et opérateurs de base nécessaires pour mettre en œuvre ces différentes fonctionnalités :

- Analyse des indices spatio-temporels pouvant être extraits par traitement informatique de la séquence d'images.
- Interface de pilotage d'une séquence vidéo et de mise en forme de la vidéo permettant l'analyse image par image de façon non linéaire ainsi que l'exportation de la séquence dans un format permettant son traitement.

4.2.2.2 Réalisation d'opérateurs d'analyse d'image :

Les recherches concernant les opérateurs de TI pertinents ont été menées indépendamment de la réalisation de l'éditeur de partitions. Les opérateurs ont été réalisés dans le but de fournir un « vocabulaire minimal » permettant de décrire, à partir de séquences vidéo, des structures de la langue des signes. Il s'agit donc d'opérateurs relativement simples et génériques exploitant en outre les conditions de réalisation du corpus (fond bleu et vêtements noirs) :

- Opérateur d'extraction de la silhouette du locuteur par soustraction des pixels appartenant au fond
- Opérateur de détection des zones "visage" et "mains" par classification des couleurs et sélection des zones ayant la couleur de la peau
- Opérateurs de suivi d'un composant défini par sa zone englobante (d'une main ou du visage) dans une version assistée par des interactions avec l'utilisateur
- Suivi des déplacements du bout du doigt
- Détection de la direction indiquée par le doigt et matérialisation de la zone pointée dans l'espace de signation.
- Détection d'indices visuels spatio-temporels
- Module de visualisation synthétique du mouvement
- Définition d'un modèle biomécanique du bras et validation d'une méthode permettant l'estimation 3D des configuration statiques du bras.



Extraits de la séquence originale (la séquence a été réalisée indépendamment du projet LSCOLIN)

Extraction de la silhouette du locuteur

Extraction des pixels appartenant à la classe « peau »

Application de l'algorithme de suivi sur la main

Exemples de signature du mouvement pour la silhouette (à gauche) pour les régions de peau

4.2.2..2.1 Etude « système »

Les recherches menées au sein du projets LS-Colin sur l'application du traitement d'images à l'étude de la langue des signes se placent dans un cadre plus large qui constitue un des axes de recherche de l'équipe TCI. En effet, les opérateurs validés dans le cadre de ce projet fournissent un premier ensemble cohérent d'outils permettant d'extraire des primitives significatives de la séquence d'images. L'exploitation de ces opérateurs au sein de systèmes d'interprétation d'images permettant de construire des descriptions de concepts de la langue des Signes (ici des configuration de gestes, des paramètres lingusitiques, etc …)(voir Dalle 2001). L'intérêt de ce système pour le projet LS-COLIN réside dans l'approche utilisée pour formuler ce que l'on recherche dans l'image : les spécifications se font par des interactions directes dans les images présentées par le système. L'utilisateur-linguiste est donc amené à définir ses objectifs en termes visuels et il voit, en retour, ce que le système a compris et sait

retrouver d'après ses indications. La spécification et l'extraction d'indices ne passent donc pas par l'intermédiaire d'un langage de requête qui risquerait de biaiser l'analyse.

4.3 Réalisation

Plusieurs étapes de réalisation de l'éditeur de partition ont été envisagées :

- un premier prototype intégrant la vidéo mais ne disposant pas des outils de traitement d'images, qui doit permettre une première évaluation sur les aspects utilisabilité (ergonomie, fonctionnalités) ;
- un deuxième prototype intégrant une partie des outils de traitements d'images, qui doit lui-même être évalué par les utilisateurs ;
- un troisième prototype intégrant tous les outils de traitements d'images envisagés, lui aussi évalué.

Dans le cadre du projet LS-Colin, seul le premier prototype a été réalisé. Il doit encore être évalué en situation d'usage réel. L'intégration des outils de traitement d'images et les évaluations successives se feront dans une étape ultérieure.

4.4 Évaluation

Avant d'être intégrés dans l'éditeur de partition, les outils de traitement d'images doivent eux-mêmes être évalués, car ils doivent être suffisamment fiables et robustes pour apporter une aide réelle à la transcription. Ce travail a été retardé par le temps nécessaire à la réalisation et à l'exportation du corpus dans un format exploitable pour le traitement d'images. Il faut également noter la dichotomie qui existe entre l'algorithme effectivement exécuté par l'ordinateur opérant dans le seul domaine de l'image et l'interprétation des résultats obtenus par un linguiste qui ignore tout du fonctionnement de cet opérateur. L'étude de la robustesse se ramène alors à l'évaluation de l'adéquation d'un opérateur donné avec la tâche pour laquelle souhaite l'utiliser qui se fera au fur et à mesure de l'exploitation de ces opérateurs par les linguistes.

D'autre part un certain nombre des opérateurs, et spécialement les opérateurs de segmentation, mis en œuvre exploitent des hypothèses à priori sur les images (notamment la présence d'un fond bleu et les vêtements unis des locuteurs). Il convient néanmoins de souligner que ces opérateurs sont très délicats à mettre en œuvre sans ces hypothèses et que leur mise au point dans le cas général reste un problème ouvert.

Enfin, la stabilité des opérateurs doit être envisagée en considérant la chaîne d'interprétation dans son ensemble (segmentation =>mesure=>identification), cette chaîne pouvant contenir des mécanismes permettant la remise en cause des résultats obtenus par l'opérateur seul et éventuellement leur correction. Il faut de plus noter que la présence d'artéfacts dans les résultats obtenus par l'application d'un opérateur seul peut permettre de détecter l'occurrence d'un évènement spatio-temporel (par exemple, si l'on suit par la couleur un objet identifié comme étant une main, et que cet objet est perdu cela signifie que la main est passée devant un autre objet de la même couleur).

4.5 Améliorations possibles

Fonctionnalités

Il serait intéressant d'établir des liaisons entre différentes parties de la transcription :

- soit des liens auxquels on peut associer une référence appartenant à la liste des références déjà rencontrées, ou une référence « à créer »,
- soit des références (spatiales, temporelles) pouvant être liées aux liens existant ou à venir.

Cela permettrait par exemple d'associer un classificateur au signe standard auquel il fait référence.

Ergonomie

Il serait intéressant de pouvoir regrouper verticalement des paramètres avec des zones de couleurs différentes ou en utilisant la transparence, afin par exemple de visualiser les zones relatives aux signes standards et celles de grande iconicité.

5 Exploitation du corpus

5.1 Première transcription manuelle du corpus LS-COLIN

Les linguistes participant au projet ont décidé de réaliser une première transcription manuelle d'un court extrait du corpus LS-COLIN, l'éditeur de partitions étant encore en phase de test. Notre but premier est de vérifier l'apport (pour la transcription et pour l'analyse) d'un tel type de corpus réalisé dans des conditions techniques d'une très bonne qualité.
Nous avons choisi de travailler à partir d'un système de transcription « en partition » commun à tous les linguistes à fin de pouvoir partager et échanger les connaissances sur la qualité du corpus ainsi que sur l'analyse des structures de la langue. Ce travail s'est avéré ainsi utile et complémentaire au premier travail de transcription (Blanche Neige) réalisé entre décembre et mars 2001.
Grâce à l'incrustation manuelle des images dans le système de partition nous avons pu, en équipe de linguistes, examiner quelques aspects structuraux de la LSF qui permettent de valider la présence de certains indices récurrents qui peuvent être pertinents lors du traitement d'images (prédiction-vérification) et lors de la segmentation de l'énoncé par un système automatique de transcription des LS.

5.1.1 Objectifs

- Examiner la qualité de l'enregistrement lors de la transcription et de l'analyse.
- Estimer un éventuel gain de temps lors de l'extraction et le traitement des images à insérer dans la partition manuelle.
- Mettre en évidence l'apport d'un tournage de qualité vidéo numérique professionnelle et l'utilisation de trois plans différents : notamment les plans visage et profondeur qui exposent de façon explicite certains indices visuels non manuels très fin (mimique faciales et mouvement du corps – par exemple déhanchement) qui ne sont pas forcément visibles lors des enregistrements classiques sur plan américain et qui sont des éléments constitutifs de la structure des LS.

5.1.2 Démarche

Suite à la numérisation en trois plans synchronisés du récit *Le Cheval*, nous avons choisi une séquence de ce récit pour pouvoir réaliser la transcription. C'est ce récit raconté par Khadra Abdelghefar [Corpus LS-COLIN] - Séquence de 06:15 [PRET] à 06:27 [personnage transféré « Cheval »].
Un premier travail d'extraction et de traitement d'images a été réalisé par une des linguistes. Ensuite, une grille (partition en paramètres) a été construite et remplie par l'ensemble des linguistes. L'extraction de l'ensemble des images a été faite sur le plan américain, néanmoins nous avons extrait certaines images sur les deux autres plan afin de voir la qualité et l'apport visuel de ces images.

5.1.2.1 Remarques sur le travail d'extraction et de traitement d'images

La numérisation de l'extrait avec les trois plans synchronisés et incrustés dans la même fenêtre permet une visualisation remarquable au niveau de l'analyse en vidéo de la langue. Néanmoins, lors de l'extraction d'image nous n'avons pas pu incruster l'image telle quelle sur la transcription car les trois plans ne donnaient pas une vue analytique de l'image. Ainsi nous avons dû utiliser un logiciel de traitement d'images pour sélectionner la portion de l'image à mettre sur le document :

- Le corps du locuteur ainsi que l'espace de signation ont été sélectionnés en soustrayant les autres plans et les parties qui ne sont pas pertinentes à l'analyse. Ce travail permet de donner une impression de zoom à l'image analysée (caméra 1).
- Même si le passage par un logiciel de traitement d'image était nécessaire, le travail de préparation de l'image à insérer sur la transcription a été simplifié car les images sont d'une très bonne qualité, donc un temps supplémentaire pour le travail de retouche n'a pas été utile (luminosité, contraste, brouiller, affiner…)
- Un travail d'incrustation de flèches sur l'image sélectionnée a été fait visant la visualisation du mouvement en question.
- Grâce au plan visage, il a été possible de sélectionner les traits pertinents du visage (la direction, le clignement du regard et les différentes mimiques faciales)
- Grâce au plan profondeur, certains mouvements ont été dégagés (qui ne sont pas entièrement visibles sur le plan américain – en deux dimensions). Ce plan nous donnent la possibilité de voir le mouvement en profondeur (utilisation de l'espace)

5.1.2.2 Résultats

La résolution de la définition de l'image nous a beaucoup facilité le travail de transcription et de dégagement de paramètres. Grâce à la qualité d'enregistrement de ce corpus, à la fin de notre travail de transcription , nous avons pu détecter avec précision :

- La pertinence du clignement des yeux en tant qu'indice de segmentation (bornes de l'énoncé)
- La structure interne des transferts, notamment celle du transfert situationnel.
- Le rôle du regard (protagoniste de l'énoncé et protagoniste de l'énonciation)

5.1.2.3 Ouverture vers les pistes de recherches

Même si nous sommes encore loin d'une reconnaissance détaillée des gestes par l'ordinateur, les paramètres détectés nous incitent à approfondir la réflexion sur l'identification de gestes non manuels (notamment la direction du regard et le clignement des yeux) en tant qu'indice de segmentation des LS.

Il est clair que lorsqu'on aura un système informatisé respectant des contraintes d'ouverture et d'évolutivité au niveau des paramètres retenus ainsi que la possibilité de personnalisation de la présentation des différentes transcriptions de corpus vidéo des LS le travail d'analyse sera moins pénible au niveau du traitement de l'image et plus rentable au niveau de l'analyse et de la mise en relief de la structure interne des LS.